

CLAIMS

What is claimed is:

1. A method for distributing data in a system that includes a plurality of servers, the method comprising:

identifying ones of the servers to store a replica of the data based on at least one of utilization of the servers, prior data distribution involving the servers, and failure correlation properties associated with the servers; and

placing the replicas of the data at the identified servers.

2. The method of claim 1, wherein the identifying ones of the servers includes:

identifying underutilized ones of the servers as candidates to store the replicas of the data.

3. The method of claim 2, wherein the underutilized servers are identified based on

disk space usage below a determined amount.

4. The method of claim 1, wherein the identifying ones of the servers includes:

identifying ones of the servers that have not been involved in a recent data distribution as candidates to store the replicas of the data.

5. The method of claim 1, wherein the identifying ones of the servers includes:

identifying system conditions that affect two or more of the servers, and

identifying ones of the servers as candidates to store the replicas of the data based on the identified system conditions.

6. The method of claim 1, wherein a number of the replicas of the data stored by the servers is user-configurable.

7. A system for distributing chunks in a network that includes a plurality of servers, comprising:

means for selecting ones of the servers to store replicas of the chunks based on at least one of utilization of the servers, prior chunk distribution involving the servers, and failure correlation properties associated with the servers; and

means for storing the replicas of the chunks at the selected servers.

8. A file system, comprising:

a plurality of servers that store replicas of chunks; and

a master connected to the servers, the master being configured to:

identify one or more of the servers to store a replica of a chunk based on at least one of utilization of the servers, prior chunk distribution involving the servers, and failure correlation properties associated with the servers, and

place the replicas of the chunk at the identified one or more servers.

9. A method for distributing chunks of data in a system that includes a plurality of servers that store replicas of the chunks, the method comprising:

monitoring total numbers of replicas of the chunks available in the system;

identifying chunks that have a total number of replicas below one or more chunk thresholds;

assigning priorities to the identified chunks including at least one of:

assigning a higher priority to one of the identified chunks whose total number of replicas is farther away from a corresponding one of the one or more chunk thresholds than another one of the identified chunks whose total number of replicas is closer to another corresponding one of the one or more chunk thresholds,

determining priorities for the identified chunks based on whether the identified chunks are associated with active files, and

determining priorities for the identified chunks based on whether the identified chunks are blocking progress within the system; and

re-replicating the identified chunks based substantially on the assigned priorities.

10. The method of claim 9, wherein the one or more chunk thresholds are user-configurable.

11. The method of claim 9, wherein the one or more chunk thresholds are the same for all chunks.

12. The method of claim 9, wherein the one or more chunk thresholds are set for each class or type of chunk.

13. The method of claim 9, wherein the assigning priorities to the identified chunks alternatively or additionally includes:

determining priorities for the identified chunks based on how close the total numbers of replicas for the identified chunks are to the one or more chunk thresholds.

14. The method of claim 9, wherein the re-replicating the identified chunks includes: cloning a higher priority one of the identified chunks prior to cloning to a lower priority one of the identified chunks.

15. The method of claim 14, wherein the cloning includes: instructing one of the servers to copy one of the identified chunks from another one of the servers.

16. The method of claim 9, wherein the re-replicating the identified chunks includes: identifying ones of the servers based on at least one of utilization of the servers, prior data distribution involving the servers, and failure correlation properties associated with the servers, and

instructing the identified servers to copy ones of the identified chunks from other ones of the servers.

17. A system for distributing data in a network that includes a plurality of servers that store replicas of the data, the system comprising:

means for monitoring total numbers of the replicas available in the network;

means for identifying data that has a total number of replicas below one or more thresholds;

means for prioritizing the data, including at least one of:

means for assigning a higher priority to data whose total number of replicas is farther away from a corresponding one of the one or more thresholds than data whose total number of replicas is closer to another corresponding one of the one or more thresholds,

means for determining priorities for the data based on whether the data is associated with active files, and

means for determining priorities for the data based on whether the data is blocking progress within the network; and

means for re-replicating the data based on the assigned priorities.

18. A file system, comprising:

a plurality of servers configured to store replicas of chunks of data; and

a master connected to the servers, the master being configured to:

monitor total numbers of valid ones of the replicas stored by the servers,

identify chunks that have a total number of valid replicas below one or more thresholds,

assign priorities to the identified chunks by at least one of:

assigning a higher priority to one of the identified chunks whose total number of valid replicas is farther away from a corresponding one of the one or more thresholds than another one of the identified chunks whose total number of valid replicas is closer to another corresponding one of the one or more chunk thresholds,

determining priorities for the identified chunks based on whether the identified chunks are associated with active files, and

determining priorities for the identified chunks based on whether the identified chunks are blocking progress within the file system, and  
re-replicate the identified chunks based substantially on the assigned priorities.

19. A method for redistributing chunks of data in a system that includes a plurality of servers that store replicas of the chunks, the method comprising:

monitoring utilization of the servers;

determining whether to redistribute any of the replicas;

selecting one or more of the replicas to redistribute based on the utilization of the servers;

selecting one or more of the servers to which to move the one or more replicas; and

moving the one or more replicas to the selected one or more servers.

20. The method of claim 19, wherein the utilization of the servers relates to an amount of free disk space available at the servers.

21. The method of claim 19, wherein the selecting one or more of the servers includes:

identifying underutilized ones of the servers as candidates to which to move the one or more replicas.

22. The method of claim 21, wherein the underutilized servers are identified based on disk space usage below a determined amount.

23. The method of claim 19, wherein the selecting one or more of the servers includes:

identifying ones of the servers that have not been involved in a recent redistribution as candidates to which to move the one or more replicas.

24. The method of claim 19, wherein the selecting one or more of the servers includes:

determining failure correlation properties associated with the servers, and

identifying ones of the servers based on the failure correlation properties as candidates to which to move the one or more replicas.

25. The method of claim 19, wherein the moving the one or more replicas includes:  
deleting the one or more replicas from one or more of the servers, and  
instructing the selected one or more servers to copy the one or more replicas from another  
one or more of the servers.

26. A system for redistributing data in a network that includes a plurality of servers  
that store replicas of the data, the system comprising:  
means for monitoring utilization of the servers;  
means for selecting one or more of the replicas to redistribute based on the utilization of  
the servers;  
means for identifying one or more of the servers to which to move the one or more  
replicas; and  
means for redistributing the one or more replicas to the identified one or more servers.

27. A file system, comprising:  
a plurality of servers configured to store replicas of chunks of data; and  
a master connected to the servers, the master being configured to:  
select one or more of the replicas to redistribute based on utilization of the  
servers,  
identify one or more of the servers to which to move the selected one or more  
replicas, and  
move the selected one or more replicas to the identified one or more servers.